DOCUMENT RESUME

ED 136 149 CG 011 189

AUTHOR Gagne', Ellen D.; And Others
TITLE Type of Post-Question and Accuracy of Concept Classification in Learning from Prose.
PUB DATE 3 Sep 76
NOTE 25p.; Paper presented at the Annual Convention of the American Psychological Association (84th, Washington, D.C., September 3-7, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Associative Learning; College Students; *Directed Reading Activity; *Learning Processes; *Mnemonics; *Reading; *Recall (Psychological); *Reinforcement; Research Projects; Study

ABSTRACT

College students (n=192) read two 1-1/2 page passages on the concepts of autism and trochaic meter. After reading, four independent groups answered questions requiring either a summary, the generation of a new example of the concept, a listing of the critical attributes of the concept, or the identification of a new example. All groups received feedback, and a control group received feedback but no post-question. Alternate forms of a test of concept classification were given one day and one week after reading. The identification post-question group showed greater accuracy in the classification of new examples of the autism concept than did the control group after one day, but no groups differed after one week. (Author)

Paper presented at APA, 1976, convention
September 3, 1976
Washington D.C.

Type of Post-Question and Accuracy of Concept Classification in Learning from Prose

Ellen D. Gagne', Sam P. Broughton, P. Denise Eggleston, Suzy E. Holmes, Priscilla P. Hawkins, and Kenneth L. Sheldon

University of Georgia

Running Head: Concept Learning from Prose

Abstract

College students (n=192) read two 1-1/2 page passages on the concepts of autism and trochaic meter. After reading, four independent groups answered questions requiring either a summary, the generation of a new example of the concept, a listing of the critical attributes of the concept, or the identification of a new example. All groups received feedback, and a control group received feedback but no post-question. Alternate forms of a test of concept classification were given one day and one week after reading. The identification post-question group showed greater accuracy in the classification of new examples of the autism concept than did the control group after one day, but no groups differed after one week.

Type of Post-Question and Accuracy of Concept Classification in Learning from Prose

Date Submitted: June 9, 1976

As Markle (1975) has recently noted, the state of the art of teaching concepts from written material is less than primitive. Typically, textbooks define a concept and give one or two examples of it. From this information the reader is expected to be able to recognize all other examples of this concept in addition to non-examples that might be confusing because of some shared attributes. Since, in numerous studies, adjunct aids such as post-questions have been found to facilitate the acquisition of verbal information from prose (of. Rothkopf, 1966, Rothkopf & Bisbiscos, 1967), the present study sought to determine whether such aids might also facilitate the acquisition of concepts from prose. By "acquisition of concepts" we mean the acquisition of the ability to classify new instances of the concept.

Two theoretical approaches to concept learning from prose which exist each lead to different expectations about the effectiveness of certain types of post-questions. The first approach, which will be referred to as the "meaningful processing" approach, claims that "procedures which induce meaningful processing facilitate learning from connected discourse" (Anderson & kulhavy, 1972, p. 390). Watts and Anderson (1971) employed this approach in their study of learning principle described in the passage to a new situation improved the learning of that principle more than did post-questions requiring

subjects either to apply the principle to a situation that was the same as the one described in the passage or to identify the name of the person associated with the principle. They concluded (p. 393) that "Answering application questions facilitates later performance by encouraging students to process the contents of instruction more thoroughly, in fact to transform it, in the effort to apply it to a new situation." Assuming that principle and concept learning have much in common (Gagne', 1974), this interpretation would lead to a prediction that any post-questions requiring subjects to transform relevant information would improve concept learning from text.

Whereas the meaningful processing approach highlights some notions about information processing that may apply across types of learning outcomes, a second approach to the learning of concepts from prose, here referred to as the "concept analysis" approach, emphasizes the unique characteristics of processing concepts. It has been suggested that concept learning involves both the ability to discriminate the relevant (defining) attributes of a concept from irrelevant attributes and the ability to generalize to instances of the concepts that have atypical irrelevant attributes (Markle & Tiemann, 1969). Studies have demonstrated the utility of contrasting examples of a concept with non-examples that share some of the relevant attributes (Houtz, Moore, & Davis, 1973; Tennyson, 1973). Apparently this procedure encourages the discrimination of the relevant attributes. In addition, it has been shown that exposing subjects to "divergent" examples that vary widely on irrelevant attributes improves concept learning (Houtz,

Moore, & Davis, 1973; Tennyson, Woolley, & Merrill, 1972). This procedure seems to discourage the association of an irrelevant attribute with the concept, hence reducing the probability of undergeneralization. Applying these findings to the learning of concepts from prose, it would seem that post-questions which increase the probability of discrimination of relevant attributes and/or generalization across irrelevant attributes would be most effective. One type of post-question that could produce both discrimination and generalization would be one asking subjects to identify a new example of the concept which had different irrelevant attributes than the example given in the passage and which was embedded in a set of nonexamples that shared some but not all of the relevant attributes of the concept.

In summary, the meaningful processing approach emphasizes transformation of information, a characteristic common to all meaningful information processing. The concept analysis approach, on the other hand, emphasizes particular kinds of transformations (discrimination and generalization) needed for concept learning. The present study is an attempt to determine which approach leads to more accurate predictions about the effectiveness of certain types of post-questions. Toward this end, subjects read brief passages which listed the defining attributes of a concept, gave an example, and gave some related historical information. The four types of post-questions used were ones asking subjects to (1) identify a new example of the concept, (2) give their own new example of the concept, (3) summarize the passage, and (4) list the distinguishing characteristics of the concept.

Both approaches would predict improved learning for subjects required to identify a new example of the concept. That is, assuming the new example and nonexamples from which the subject chose had the characteristics described above, one would argue from a concept analysis approach that this question would encourage the discrimination of defining attributes and generalization across irrelevant attributes. One would argue from a meaningful processing approach that an identify type question requires subjects to think carefully about what they have read and so to transform it to a semantic level as they apply it in a new situation.

Similarly, for the question requiring subjects to give a new example of the concept, the meaningful processing approach would predict improved learning because information in the passage must be processed at a semantic level if the subject is to generate a new example. For this question, however, the concept analysis approach would not predict facilitation since the post-question fails to provide contrasting stimuli that encourage discrimination. The subject might generate a new example that did not have all the defining attributes and might even receive feedback that the example is incorrect, but this feedback does not necessarily encourage attention to defining attributes since feedback can be received passively. In addition, there is no guarantee that the new example generated by the subject will have divergent irrelevant attributes. In fact, it is more likely that it will not since the subject will probably use the example in the passage as a model. Hence, it is unlikely that the generalization process necessary for concept learning will occur

The post-questions requiring subjects to either list the distinguishing characteristics of the concept or summarize the passage were included because it is our impression that these types of questions are frequently posed by teachers. Despite this fact, neither theoretical approach would predict an improvement in learning for either type of question. For a summary question, the subject must transform information while selecting and condensing, but the information selected may not be relevant to the concept to be learned. Hence its transformation will not improve the learning of that concept. By contrast, the list post-question requires little transformation of information since the defining attributes are essentially listed in the passage. Under such conditions, neither theoretical position regarding concept learning from prose would predict improved concept learning.

Method

Subjects

Subjects were 192 students (32 in each of six treatment groups) from introductory sociology classes at a large southeastern university. Participation was encouraged but not required, and no course credit was given for it. Only two students in the classes visited did not participate. The average verbal SAT score for the entire sample was 459. A one-way analysis of variance revealed no significant differences between treatment groups with respect to this score, F (5, 161) = .79, p = .558.

Materials

Concepts. Two concepts were used--autism and trochaic meter. For the purposes of this study, autism was defined by the following attributes: (1) onset age of the problem (before 30 months), (2) slow

or abnormal speech development, (3) abnormal interpersonal relationships, and (4) repetitive or ritualistic behaviors. Trochaic meter had two defining attributes: (1) two syllables per foot of poetry and (2) stressed syllable precedes unstressed syllable (Tennyson, Woolley, & Merrill, 1972). Two 1-1/2 page passages, each describing one of the concepts, were constructed to mimic "typical" textbook writing. Each contained some background information, a description of the relevant attributes of the concept, and an example of the concept. The proportion of the concept-relevant to total number of sentences was .25 for the trochaic meter passage and .45 for the autism passage. The two passages were presented together in a booklet with passage order counterbalanced across treatment groups.

Post-questions. A separate sheet containing a post-question was inserted after each passage, the type of post-question varying with the treatment group. The Give group was asked to give a new example of the concept presented in the passage they had just read; the Summarize group was asked to summarize the passage; the List group was asked to list the distinguishing characteristics of the concept; and the Identify group was asked to identify a new example of the concept from a set including one example and three nonexamples. The example was divergent from the example given in the passage in that it had different irrelevant attributes. For instance, the example of autism given in the passage was of a boy born prematurely while the example given in the identify post-question was of a girl born normally. All of the nonexamples for autism were "close-in" (Markle, 1975) in that each had manifestations of all but one of the

relevant attributes of the concept. For instance, for autism, one of the nonexamples was a child who had slow speech development, abnormal interpersonal relationships and these symptoms had been detected when he was less than a year old. The child did not, however, have any repetitive or ritualistic behaviors.

Feedback. In addition to the four post-question groups, two control groups received no post-questions. Following the post-question page was a feedback page which for all groups contained a listing of the relevant attributes of the concept but which was adapted appropriately to each type of question. For example, the feedback for the Summary group started out by stating, "Your summary was correct if it included at least the following points:..." and then listed the relevant attributes, while the feedback for the Identify group started out by stating "If you chose (b) you are correct because (b) includes all the following characteristics:..." and then listed the relevant attributes. Feedback was included to increase the external validity of the findings for situations in which textbooks or teachers provide both questions and feedback. One control group received no post-question and no feedback (Control) and one received feedback only (Control with Feedback).

Test. A two-item test with one item per concept, was used to assess learning. For each item, subjects were asked to identify all new examples of the concept from a set of five possibilities. In addition, students were asked to rank their confidence in the correctness of their choices on a scale of 1 (very confident) to 5 (not very confident). Two forms of the test, each containing different examples, were constructed so that

concept acquisition could be measured at two different times, with the form each subject took at each time being counterbalanced. Two of the five examples were correct for each item on both forms of the test, one having irrelevant attributes similar to the example given in the passage and one having dissimilar irrelevant attributes.

Procedure

During regular sociology class time, booklets were distributed to students in a predetermined random order. Students were instructed to read the passages carefully and were told that they would be tested over the material in the passages sometime in the future. Students who received post-questions answered them in writing and could refer back to the passage while answering them and while reading the feedback. Students recorded the time when they started to read and the time when they completed reading the last feedback page. The next day (Day 1) students took their first test and a week later (Week 1) they took an alternate form of the test. Students worked at their own speed both while reading the passages and while completing the tests.

Results

Categories of Concept Identification. Each student's responses to the test were categorized as either accurate, overgeneralization, undergeneralization, or misconception. An accurate response was counted if the student identified both correct examples and none of the nonexamples; an overgeneralization was counted if the student identified both correct examples but also identified one or more nonexamples as being correct; an undergeneralization was counted if the student identified only one of the two correct examples and none of the nonexamples; a misconception was

counted if the student identified as correct only one of the correct examples and also one or more of the nonexamples as correct. The chance level for accuracy, overgeneralization, undergeneralization, and misconception, respectively, was .03, .09, .06, and .82. Table 1 indicates the proportion of subjects whose responses fell into each of these categories by group and concept for Day 1. Forty-seven percent of the students in the Identify goups were accurate on the autism concept as compared to 22% and 25% of the control groups, and 28% of the next highest post-question group, the List group. A chi-square test of association was used to determine if the proportion of subjects in the accurate category and a category consisting of the sum of the other three categories (inaccurate) differed for the Identify and Control with Feedback groups. The obtained $\chi^2 = 4.44$; df = 1, was significant ($p < .05$). Similar comparisons were conducted for each control group with every other post-question group. No other significant differences were obtained.

Insert Table 1 about here

For the concept of trochaic meter, the Identify group was again the most accurate (16% vs. 9% for the next highest groups). However, no significant associations between groups and patterns of accuracy were found, probably because of the low level of accuracy in general. It is interesting that within the inaccurate responses, the Identify group had more undergeneralizations (44% vs. the next highest group which had only 13%) and fewer misconceptions (37% vs. the next lowest group which had 66%) than any other group. A test of association between group (Identify

vs. Control with Feedback) and category of response (Undergeneralisation vs. all others) was significant, $\chi^2 = 7.74$; df=1; $p < .01$. Further exploration of the undergeneralizers in the Identify group revealed that in every case the student correctly identified the test example that had irrelevant attributes similar to the passage example but failed to identify the test example that had dissimilar irrelevant attributes.

Although the pattern of results was in most cases similar at Week 1 with the Identify group still being the most accurate (see Table 2), no significant differences were obtained.

Insert Table 2 about here

Associations Between Verbal SAT Score and Accuracy. If adjunct aids are effective they should reduce the degree of association between general verbal ability and accuracy. Table 3 shows the increase in probability of predicting verbal ability (above or below the group median) given knowledge of the student's response pattern (accurate or inaccurate) for the Day 1 and Week 1 tests (lambda index of predictive association, Hays, 1963). The significance of the association between verbal SAT score and accuracy was tested using a chi-square test. For Day 1, there were no significant associations, although for the Control group the chi-square statistic was between the .10 and .05 probability levels, $\chi^2 = 3.38$, df = 1. For Week 1 the only significant association between verbal ability and accuracy was that obtained for the Identify group.

Insert Table 3 about here

Time. The average time required by each group to read both passages and answer the questions is shown in Table 4. The groups, ranked from longest to shortest average time, were Summarize, Give, List, Identify, Control with Feedback, Control. A one-way analysis of variance was conducted to test for differences between groups in time required to complete the tasks. A significant effect due to groups was obtained, F (5, 191) = 71.59, $p < .01$. A Scheffé's test applied to the group means indicated that the Summarize and Give groups were each significantly different from every other group, and the List group was significantly different from every other group except the Identify group. The two control groups did not differ significantly from each other, nor did the Control with Feedback group differ significantly from the Identify group. Thus, it would appear that the greater accuracy demonstrated by the Identify group is not due to a greater amount of time spent on the passage.

Insert Table 4 about here

Practice effect. It is possible that practice with a certain type of post-question for the first concept increases accuracy in concept classification for the second concept encountered. To test this possibility, we obtained the frequencies of subjects who were accurate versus inaccurate on the Day 1 test for the first and second concept studied. Chi-square tests revealed no significant association of accuracy with concept order. A similar test in which all four post-question groups were pooled was also not significant.

Discussion

The data support a concept analysis approach to learning concepts from text. The only group that showed improved learning was the Identify group. We think that this improvement in the learning of the concept autism was a result of both better discrimination of relevant attributes and greater generalization a[illegible]elevant attributes caused by the comparisons that could be mad[illegible]tween examples and nonexamples. Specifically, for the Identify post-question, in considering case (a), a nonexample, and case (b), an example, the student would recognize that there was only one critical difference between them which was the age of onset of the problem. Similar comparisons between the example and the other two nonexamples would have brought attention to two of the other defining attributes. Once case (b) was recognized as the correct example, the student would also recognize that certain qualities such as sex and normalcy of birth are irrelevant to the concept since they varied between the passage example and the post-question example. An alternative explanation of the effectiveness of an identify type post-question is that there was positive transfer of test-taking skills from the identify post-question to the test, which also required identification of new instances. A study is presently being conducted to test this possibility.

The results for trochaic meter suggest a limiting condition for the effectiveness of identification post-questions. The difficulty of that concept was quite possibly due to students not understanding the prerequisite concept of stress and hence not being able to recognize stressed and unstressed syllables. Support for this possibility comes from the finding of Merrill and Tennyson (1971) that providing students with an

explanation of the statement "stressed syllable followed by an unstressed syllable" improved the learning of trochaic meter from a definition and set of examples.

Another interesting finding for trochaic meter was that the high frequency of undergeneralizers in the Identify group was entirely due to students who identified the correct similar example but failed to recognize the correct dissimilar example. The dissimilar examples used on the test were not only dissimilar on irrelevant attributes but also were examples that, unlike those used in the passage and post-questions, had a very low probability of being identified as correct (Tennyson, Woolley, & Merrill, 1972). Low probability examples have attribute values that apparently are difficult to recognize. It would seem then that the Identify group students were able to discriminate clear nonexamples and examples but failed when it came to the identification of a less clear example. A similar result has been obtained by Tennyson (1973). Research should be conducted to determine what type of adjunct aids might alleviate this problem of not recognizing the full range of relevant attribute values.

The finding of an association between verbal SAT and accuracy on the Week 1 test for the Identify group is intriguing. While it may be a chance result, another possibility is that verbal SAT correlates with the type of memory system a person uses given an Identify post-question. Persons scoring high may integrate the concept they have induced from the examples and nonexamples into their semantic memory system, while persons scoring low may simply keep an episodic record of the post-question event. If it is true that semantic memory is more permanent than episodic memory (Tulving, 1972) then one would expect both systems to be effective for

Day 1 retention, but only the semantic system to be effective by Week 1.

Unlike many studies in the area of learning from prose, the present one attempted to maintain a situation as close to a natural study situation as possible by allowing students to refer back to the passage while answering post-questions and by providing feedback. Thus, we feel justified in saying that the practical implication of the present results is

fication questions which use dissimilar examples and close-in nonexamples can facilitate concept learning from prose without substantially increasing inspection time. Apparently facilitation only occurs if the relevant attributes themselves are already recognizable. There should be replications of this experiment over a wide variety of concepts and populations.

Reference Note

1. Merrill, M. D. & Tennyson, R. D. Attribute prompting variables in learning classroom concepts. Provo, Utah: Brigham Young Univ., Division of Communication Services, Working Paper No. 28, 1971.

References

Anderson, R. C., & Kulhavy, R. W. Learning concepts from definitions. American Educational Research Journal, 1972, 9, 385-390.

Gagne', R. M. Essentials of Learning for Instruction. Hinsdale, Illinois: The Dryden Press, 1974.

Hays, William L. Statistics. New York: Holt, Rinehart, & Winston, 1963.

Houtz, J. C., Moore, J. W., & Davis, J. K. Effects of different types of positive and negative instances in learning "nondimensioned" concepts. Journal of Educational Psychology, 1973, 64, 206-211.

Markle, S. M. They teach concepts don't they? Educational Researcher, 1975 (June), 4, 3-9.

Markle, S. M. & Tiemann, P. W. Really understanding concepts. Champaign, Ill.: Stipes, 1969.

Rothkopf, E. Z. Learning from written materials: an exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 1966, 3, 241-249.

Rothkopf, E. Z. & Bisbiscos, E. E. Selective facilitative effects of interspersed questions on learning from written materials. Journal of Educational Psychology, 1967, 58, 56-61.

Tennyson, R. D. Effect of negative instances in concept acquisition using a verbal learning task. Journal of Educational Psychology, 1973, 64, 247-260.

Tennyson, R. D., Wooley, F. R., & Merrill, M. D. Exemplar and nonexemplar variables which produce correct concept classification behavior and specified classification errors. Journal of Educational Psychology, 1972, 63, 144-152.

Tulving, E. Episodic and semantic memory. In Tulving, E. & Donaldson, U. (Eds.), Organization of memory. New York: Academic Press, 1972.

Watts, G. H. & Anderson, R. C. Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 1971, 62, 387-394.

Footnote

[1]Requests for reprints should be sent to Ellen D. Gagne', Department of Educational Psychology, University of Georgia, Athens, Georgia 30601.

Table 1

Proportions of Subjects in Each Response Category for Each Concept Study 1

(A=Accurate, O=Overgeneralization, U=Undergeneralization, and M=Misconception)

	Autism				Trochaic Meter			
Group	A	O	U	M	A	O	U	M
Identify	.47	.12	.16	.25	.16	.03	.44	.37
Give	.22	.25	.12	.41	.09	.13	.12	.66
List	.28	.22	.06	.44	.06	.06	.03	.85
Summary	.16	.28	.19	.37	.06	.06	.10	.78
Control with Feedback	.22	.22	.25	.31	.09	.12	.13	.66
Control	.25	.12	.19	.44	.09	.06	.13	.72

Tab

Proportion of [illegible] in Each Response Category for Each Concept at Week 1

	Autism				Trochaic Meter			
Group	A	O	U	M	A	O	U	M
Identify	.34	.12	.16	.38	.13	.06	.16	.65
Give	.22	.37	.22	.19	.03	.06	.16	.75
List	.19	.25	.28	.28	.03	.06	.06	.85
Summary	.22	.25	.16	.34	.03	.00	.09	.88
Control with Feedback	.22	.25	.19	.34	.03	.16	.16	.65
Control	.25	.25	.22	.28	.12	.06	.16	.66

Table 3

The Increase in Probability of Correctly Predicting Verbal SAT (Above or Below Median) Given Knowledge of Response Pattern (Accurate versus Inaccurate) on Test

Group	Day 1 Test	Week 1 Test
Identify	.07	.50*
Give	.00	.00
List	.15	.00
Summary	.07	.07
Control with Feedback	.00	.14
Control	.29	.07

* A chi-square test of association is significant at $p < .05$.

Table 4

Means and Standard Deviations for Time to Read and Answer Questions

	Identify	Give	List	Summary	Control w/ Fdbk.	Control
M	8.41	12.47	9.94	15.28	6.53	5.41
SD	1.41	3.16	2.88	3.51	1.80	1.07